**Project - ML with Python on building data**
**Data Cleaning, Exploratory Data Analysis, Model Building and Tuning**

**(30 points)**

**Due before - April 17, 2023**

Preliminaries (Data Modeling) - Develop and RDF diagram that captures the concepts, the attributes and the relationships in the dataset at Real Property Information | Real Property Information | Open Baltimore (baltimorecity.gov). - You are free to seek feedback from Seila on it. You can use the Github and .ttl files already there for definition and for the IRI of each concept and relationship. (See tutorial about RDF for more on IRI)
**Deliverable**: Show your graph.

1. Download a data file (Real Property Information | Real Property Information | Open Baltimore (baltimorecity.gov))
2. **(10 points)** Use the DataFrame API to visualize your data in Jupyter Notebook at Welcome To Colaboratory - Colaboratory (google.com). Use this Machine Learning Tutorial for visualization, machine learning, etc. Link: https://machinelearningmastery.com/python-machine-learning-mini-course/.
   a) Show measures of central tendency (You may want to consider converting categorical/nominal values into numerical values) For example, convert low, medium, high values into 1, 2, 3, respectively.
   b) Show diagrams, histograms (nominal values)
   c) Show box plots for each numerical features or for selected features
   d) Identify any outliers
   e) Plot the correlation between each pair of numerical features
3. **(10 points)** Use the steps at Pythonic Data Cleaning With Pandas and NumPy – Real Python to clean it
   a) May have to remove records with missing values or records that you think are inconsistent.
      Chart visualization — pandas 1.5.1 documentation (pydata.org)
      Plot With Pandas: Python Data Visualization for Beginners – Real Python

1. **(10 points)** Develop Python Code to do machine learning on your data set.
2. Step 1- Pick one feature of your choosing (i.e., column name to be your target)
3. Step 2 - Develop code that trains, tests and tunes a model to make accurate predictions on the target. Research examples to develop and test your own classifier.
4. Choose another column. Develop a ML model that will predict that column.

To start, just follow the tutorial. Use your own Building data, instead of the Iris dataset that they have. Here is a good tutorial on machine learning in Python. You may use the same or similar steps for your data sets. Consider using other Python machine learning APIs:

https://machinelearningmastery.com/python-machine-learning-mini-course/.

Show a model and the accuracy of the model . For example, ==you will want to be able to say something like==; My model can predict the "neighborhood" of a building with 96% accuracy.